ExG Signal Feature Selection Using Hyperdimensional Computing Encoding

Una Pale*, Tomas Teijeiro*†, David Atienza*
*Embedded Systems Laboratory (ESL), École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
†Department of Mathematics, University of the Basque Country (UPV/EHU), Spain
{una.pale, david.atienza}@epfl.ch, tomas.teijeiro@ehu.eus

***Abstract*—Wearable IoT devices and novel continuous monitoring algorithms are essential components of the healthcare transition from reactive interventions focused on symptom treatment to more proactive prevention, from one-size-fits-all to personalized medicine, and from centralized to distributed paradigms. HyperDimensional Computing (HDC) is an emerging ML paradigm inspired by neuroscience research with various aspects interesting for IoT devices and biomedical applications. In this work, we explore five HD vector encoding strategies of spatio-temporal ExG data, such as that of electroencephalogram (EEG), and test it on a use case of epileptic seizure detection. We discuss the impact of these strategies' performance, memory overhead, and computational complexity. Furthermore, we demonstrate how feature selection via the HDC framework can be accomplished by choosing a proper encoding, and results in up to 70% reduction in used features while improving performance up to 7%.**

I. Introduction

Hyperdimensional Computing (HDC) is an emerging ML paradigm inspired by neuroscience research, based on data representation in the shape of high-dimensional *hypervectors*. This paradigm shift in data representation brings various advantages in learning and hardware implementation. From a learning perspective, it opens new paths for semi-supervised, distributed, continuous online learning, or multi-centroid learning. In terms of hardware, parallelization possibilities open the way for designing efficient accelerators or in-memory computations [1]. Its lower energy and memory requirements [2], [3] enable learning on low power wearables and IoT systems. HD computing has attracted a great deal of attention for biomedical applications, especially with ExG data. It has been tested for electromyogram (EMG) gesture recognition, detection of EEG error-related potentials, recognition of emotions from electrocardiogram (ECG) and electroencephalogram (EEG), and epilepsy detection via EEG [3], [4] among others.

ExG data (e.g., EEG and EMG) are spatio-temporal, noisy and non-stationary data that require careful encoding to HD vectors to enable high-quality HD learning. However, the encoding of channel and feature information has not yet been systematically explored. Most existing literature that uses EMG or EEG data has encoded only raw data or Local Binary Patterns (LBPs) [5] to HD vectors. However, similar to standard ML approaches, the possibility of adding more features can significantly improve the power of models [6]. Thus, in this paper, we discuss possibilities for encoding information about channels and features, to capture all the relevant aspects.

As we demonstrate in this work, encoding can also enable analysis of the performance, correlations, and learning capabilities of individual features. Furthermore, feature selection is a crucial step in wearable applications, needed to remove noisy and non-informative features while also leading to more lightweight models. To the best of our knowledge, so far in the literature a clear and straightforward methodology for feature selection via HDC has not been presented. We demonstrate a 70% reduction in the features used while gaining up to 7% in terms of performance.

In this paper, we assess several approaches in the context of epileptic seizure detection. Epileptic seizures are a chronic neurological disorder characterized by the unpredictable occurrence of seizures that affects a significant portion of the world's population (0.6 to 0.8%). Due to its high inter-patient variability and unpredictable nature, it still poses open research questions. Despite pharmacological treatments, one third of patients still suffer from seizures and are subject to serious health risks and many restrictions in daily life. At the same time, wearable devices for the prediction, detection, or continuous monitoring of epilepsy in outpatient settings are not yet widely available. HD computing's unique properties make it suitable for this domain, but only with further algorithmic improvements can it reach the performance of state-of-the-art algorithms [7]. Therefore, further exploration of how to optimize HD computing and encoding for better performance is of great interest for the detection of epilepsy.

II. Related Work

HD computing has been applied so far in multiple domains where spatio-temporal data such as EEG, iEEG [2], [3] or EMG [8] is utilized. Most works utilize raw signal values or short LBPs to describe and map temporal signal changes to HD vectors [3], [4], [8]. Some of the work also utilized vector permutation to encode time information between neighboring samples. Furthermore, in [1], the authors propose energy-efficient in-memory encoding for this way of encoding spatio-temporal signals. This is an interesting approach in its simplicity, as it uses raw signal values directly. However, it is limited to only mapping one type of information to HD vectors (i.e., raw signal amplitude and/or signal change trends).

This work has been partially supported by the ML-Edge Swiss National Science Foundation (NSF) Research project (GA No. 200020182009/1), and the PEDESITE Swiss NSF Sinergia project (GA No. SCRSII5 193813/1). T. Teijeiro is supported by a Maria Zambrano fellowship (MAZAM21/29) from the University of Basque Country and the Spanish Ministry of Universities, funded by the European Union-Next-GenerationEU.

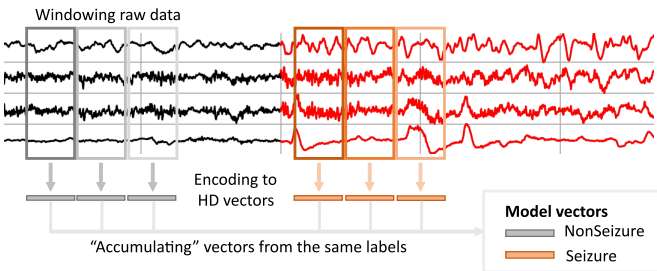

Fig. 1. Simple illustration showing the HD computing workflow and learning process from windows of seizure and non-seizure data.

In [6], the authors test how different feature types regularly used in non-HDC ML models for epilepsy detection perform in the HDC framework. They show that, indeed, the use of a variety of statistical, time, or frequency features outperforms simple raw signal encoding. Further, in [2], the authors extend their previous work [4] by adding the mean amplitude and line length features to LPB values. They resolved the problem of encoding more features to HD vectors by utilizing three independent classifiers, each with its own model vectors, where predictions are merged using a linear layer. In [9], the authors explored epileptic seizure detection using power-spectral features from iEEG data, and encountered similar problems as we point out here. They explored three different encoding approaches: 1) concatenating feature vectors to generate long HD vectors, 2) using multiple classifiers (one for each feature) and then integrating their predictions by majority voting, and 3) training one classifier using all features. In the last approach, vectors representing features and their values are first combined and then merged with the channel information. These approaches are highly interesting due to the broad spectrum of feature characteristics that are integrated. However, unfortunately, in [9] approaches were not systematically assessed across performance, memory, or complexity metrics.

III. Methods

A. Encoding spatio-temporal data for HD computing

Typical spatio-temporal (ST) data, e.g. EMG or EEG, consist of several channels positioned in different physical locations and recorded during time. This results in a 3D data structure containing information about features, channels, and time. Within the typical HDC encoding workflow, this means that initial vectors representing each of these entities must be defined, i.e. HD vectors representing each feature (F_{ID}), each possible value of features (FCh_V), and each channel (Ch_{ID}) must be generated. The question addressed in this work is how to encode all this information into HD vectors. We evaluate potential methods from several perspectives, including classification performance, memory overhead, and computational complexity. Therefore, in Fig. 2, we illustrate the different possibilities considered in this paper to encode one time window of data to an HD vector.

We use the most common type of HD vectors; binary vectors (containing only 0 or 1). Three basic arithmetic operations that are used are: 1) bundling (bitwise summation), 2) binding (bitwise XOR), and 3) thresholding to binarize vectors after summation. The bundling operation results in a vector that is, with high probability, very similar to summed vectors, while on the other hand, binding leads to a vector orthogonal to the bound vectors. Thresholding returns the final vector to binary form after summation. A typical approach in the literature dealing with simpler, non-spatial data is to bind feature vectors with feature value vectors and then bundle (and threshold) them. Translating this approach to ST data leads to two options: 1) $F \times V$ and 2) $ChFComb \times V$, where F represents features, V values of features, and $ChFComb$ features of specific channels. The difference between these approaches is that, as illustrated in Fig. 2, $F \times V$ does not include information about channels but simply bundles features (F_{ID}) and feature values from each channel (FCh_V). On the other hand, the $ChFComb \times V$ approach treats each feature and channel combination as an individual feature and gives it an independent HD vector (FCh_{ID}). This approach distinguishes between channels as formulated in (2), but can lead to large memory maps to store all initial HD vectors, especially when the dataset consists of many channels (such as iEEG data) or when many features are extracted. As this can be problematic from a memory perspective, we propose more EEG-inspired approaches: 3) $F \times Ch \times V$ and 4) $Ch \times F \times V$. Both approaches first initialize vectors for each channel (Ch_{ID}), each feature (F_{ID}), and possible feature values (FCh_V). However, they differ in the order of bundling information. $F \times Ch \times V$, as formulated by (3), bundles channels (Ch_{ID}) and feature values on those channels (FCh_V) first, followed by bundling with feature IDs (F_{ID}). In $Ch \times F \times V$ the order of bundling is inverted, as formulated in (4).

$$F \times V = \left\lfloor \sum_{i=1}^{nCh*nF} F_{IDi} \oplus FCh_{Vi} \right\rfloor \tag{1}$$

$$ChFComb \times V = \left\lfloor \sum_{i=1}^{nCh*nF} FCh_{IDi} \oplus FCh_{Vi} \right\rfloor \tag{2}$$

$$F \times Ch \times V = \left\lfloor \sum_{i=1}^{nF} F_{IDi} \oplus \left\lfloor \sum_{j=1}^{nCh} Ch_{IDj} \oplus FCh_{Vij} \right\rfloor \right\rfloor \tag{3}$$

$$Ch \times F \times V = \left\lfloor \sum_{i=1}^{nCh} Ch_{IDi} \oplus \left\lfloor \sum_{j=1}^{nF} F_{IDj} \oplus FCh_{Vij} \right\rfloor \right\rfloor \tag{4}$$

$$FA = \left\lfloor \sum_{j=1}^{nCh} Ch_{IDj} \oplus F_1 Ch_{Vj} \right\rfloor ... \left\lfloor \sum_{j=1}^{nCh} Ch_{IDj} \oplus F_{nF} Ch_{Vj} \right\rfloor \tag{5}$$

The last approach, called Feature Appending (FA), is designed to make it easier to determine how different features contribute to encoded vectors. In this approach, as illustrated in Fig. 2, channels and feature values are bound and bundled to n encoded subvector for each feature, but instead of binding it with other feature subvectors as in $F \times Ch \times V$ these vectors are appended one after another, as formulated in (5). This encoding organization enables analysis of individual subvectors. For example, in this paper, we analyze the class separability and predictive power of individual features, as well as their confidence in time with respect to other features.

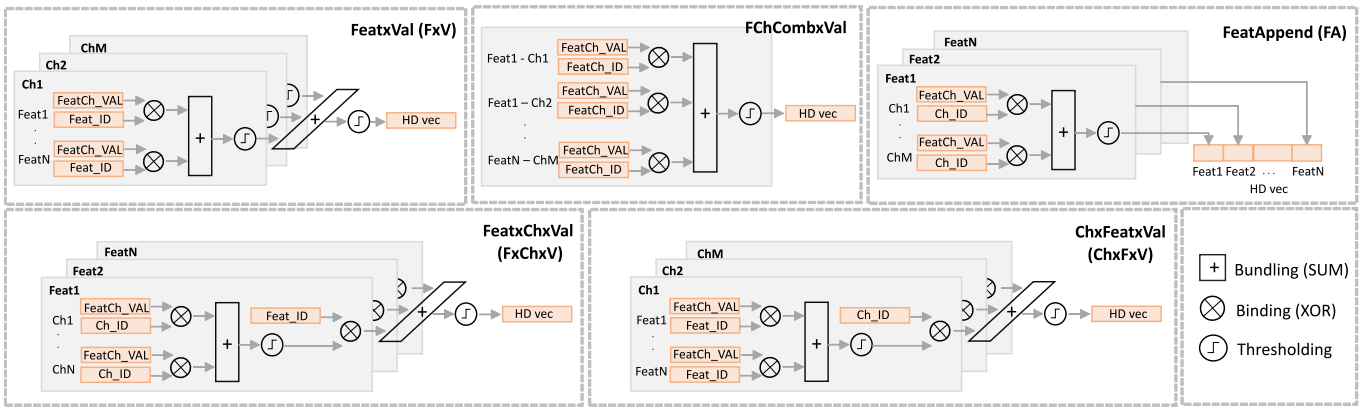

Fig. 2. Schematic of different possibilities to encode information about channels, features, and their values to HD vectors. Five different encoding approaches tested in the paper are illustrated.

B. Feature selection

In the FA approach, it is known which part of the final encoded vector comes from each feature, allowing analysis of individual features. In this paper, we define and measure several metrics for each feature:

- *Prediction*: Using only the $d = D/nF$ bits corresponding to the feature of interest, we can determine the prediction per feature as the label of the most similar class vector. D is the total hypervector length, nF the number of features, and the similarity metric is the hamming distance.
- *Feature certainty*: The certainty of each feature's label based on distances from class vectors can be quantified for each time moment. It is defined as the difference between distances from the two classes divided by the average absolute distance for all features at the same time.
- *Correlation*: Based on each feature's predictions in time, the correlation between different feature's predictions can be measured and utilized later for feature selection.
- *Class separability*: Feature separability is defined as the relative hamming distance between class vectors when using only the bits of the corresponding feature.

These metrics can be then used to perform feature selection, using three selection strategies. Each approach starts by ordering features based on specific quality metrics:

- *Feature Selection By Performance (SBP)*: On the basis of the predictions for each sample and the true labels, we can measure each feature's performance. The exact performance metric of choice can depend on the application; for epilepsy detection we define these in Sect. IV-D2.
- *Feature Selection By Confidence (SBC)*: Based on the certainty values and predictions per feature, the features can be ordered based on the highest confidence. It is calculated as how much more certain is the feature during correct predictions versus wrong predictions.
- *One-By-One (OBO) feature selection*: If the selection is made based only on feature performance, it might lead to selecting features that perform well, but are highly correlated, and thus possibly redundant. In this approach, we select features one-by-one by evaluating in each step how the performance changes when adding one of the not yet used features.

After features are ordered, the performance of the train and test set is assessed while increasing the number of features until all features are included. Prediction when using n features is given by summing the distances from the seizure and non-seizure model vectors of the individual features. From the performance curve of the training set, the optimal number of features is chosen as the minimum number of features giving maximal performance. In the end, performance on the test set is measured for the chosen reduced set of features and compared with the initial performance without feature selection.

IV. Experimental Setup

A. Dataset

We use the CHB-MIT database [10]. It consists of 24 subjects and 183 seizures, with an average of 7.6 $\pm$ 5.8 seizures per subject. 18 EEG channels that are common to all patients are used. Even if the common approach in the literature is using balanced data preparation, it can lead to highly overestimated performance [11]. Training on the entire dataset using HD computing is not feasible due to its complexity, thus, as proposed in [7], we use a data selection approach that contains all seizure segments and ten times more non-seizure data. Data is arranged in such a way that for each seizure file, seizure data is extracted and surrounded by non-seizure data randomly selected from one of the files not containing any seizure.

B. Feature Choices

We use the mean amplitude and both relative and absolute values of power spectral density in the five common brain wave frequency bands: delta: [0.5-4] Hz, theta: [4-8] Hz, alpha: [8-12] Hz, beta: [12-30] Hz, gamma: [30-45] Hz, and low-frequency components: [0-0.5] Hz and [0.1-0.5] Hz. We also included the line length feature [12] showing high discriminative power. In total, we extract 19 features from data segmented into 4-second windows with a 0.5-second step. Before extracting the features, the data is filtered with a 4th-order, zero-phase Butterworth bandpass filter between [1, 20] Hz.

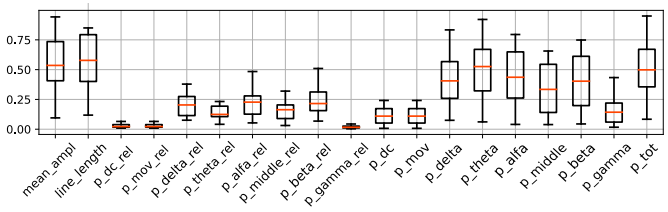

Fig. 3. Jensen-Shannon divergence of features.

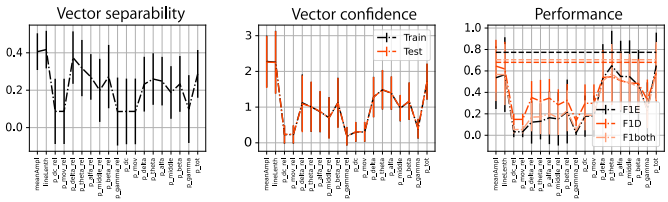

Fig. 4. Comparison of features based on FA approach. Order of features is the same as in Fig. 3. The averaged values over all subjects are shown.

C. HD Learning Workflow

The standard HDC workflow consists of a single-pass training phase, where HD vectors representing different data windows from the same label class are bundled together to form a model HD vector representing that class (illustrated in Fig. 1). This approach is simple and fast. However, all data windows are equally important, which can lead to the domination of more common patterns in the final vectors, a problem especially prevalent in highly imbalanced datasets such as those of epilepsy. As shown in [7], this situation leads to an under-representation of less common patterns and ultimately lowers performance. OnlineHD was proposed in [13] as an alternative, where each window vector is multiplied by a weight before being added to the model vector. The weight is defined by the similarity of the current vector to the current prototype vectors; the higher the similarity, the lower the weight, which helps to identify repeating patterns and mitigates model saturation. In [7] standard and OnlineHD have been compared for the use case of epileptic seizure detection, and OnlineHD was shown to have higher performance, so we use it in this work as well.

D. Validation

1) Feature comparison: A standard approach to compare features individually is the Jensen-Shannon divergence, which analyse the distance between the distributions of feature values for different classes. Furthermore, as explained in Sect. III-B, the FA appending approach enables various feature comparison metrics, namely: predictions and performance, confidence, correlation, and class separability per feature.

2) Performance evaluation: Training and evaluation are done using leave-one-seizure out cross-validation independently for each subject due to the subject-specific nature of epileptic seizures. In the end, we report performance as average over all subjects. The performance of the classifier is evaluated with respect to seizure episode detection and duration-based detection, measuring sensitivity, predictivity, and F1 score. The performance at the episode level groups the signal into blocks of seizure and non-seizure. The performance at the duration level, cares about the correct prediction of each sample, meaning that seizures need to be predicted correctly over their entire duration. Finally, we combine the F1 scores for episodes ($F1E$ and $F1D$) using the geometric mean as $F1DEgmean$. In epilepsy detection, raw label predictions often lead to unrealistic behavior of seizure dynamics (e.g., seizures lasting only a few seconds or seizures that are only a few seconds apart). Thus, label post-processing is an integral part of the pipeline, and we post-process raw label predictions by performing a moving average with majority voting, using a window size of 5s.

V. Results

A. Feature comparison

Fig. 3 shows the Jensen-Shannon divergence of the 19 features we used and their distribution over all channels for all subjects. There is a clear difference between features, where the mean amplitude, line length, total energy, and absolute spectral powers of the delta, theta, alpha, beta, and middle-range are quite discriminative. Relative powers seem to be less discriminative than absolute values. Fig. 4 compares features based on measures extracted using the FA approach. More specifically, the separability of the vectors for the two classes, the average confidence, and the performance of each feature are shown. In the performance subplot, the horizontal line represents the performance achieved when using all the features, and shows that no single feature reaches the performance of all features, but some of them are quite close (i.e., mean amplitude, line length, total power, and power of delta and theta). These results confirm that the FA approach can indeed be used to investigate different properties of individual features.

B. Encoding comparison

Next, we compare different encoding approaches. Fig. 5 shows the average performance for the episode and duration metric, for all subjects without any post-processing. $F \times V$ encoding, which does not account for channel information, results in lower performance than strategies that account for channel information. There is no significant performance difference between the three approaches that include channel information: $F \times Ch \times V$, $Ch \times F \times V$ and $ChFComb \times V$. The FA approach, even distinguishing between channels, yields a lower performance than the three approaches that utilize channel information, probably due to the smaller number of dimensions per feature. Yet, FA still outperforms the $F \times V$ approach. Fig. 6 shows the memory required (as a relative ratio) to store all HD vectors (for channels, features, and values) for each approach. Due to the large number of combinations of features and channels, the $ChFComb \times V$ approach is the most memory-demanding one. The FA approach requires the least amount of memory, as the base vectors have lower D/nF dimensions. Furthermore, the relative number of operations needed to encode one window of data is shown as well in Fig. 6. The FA approach requires significantly less operations due to the smaller number of dimensions per vector for each feature.

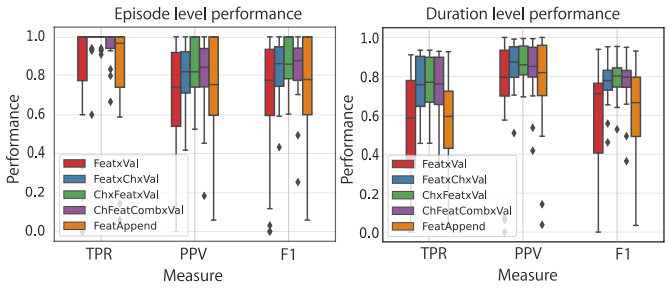

Fig. 5. Performance of different EEG encoding approaches. True Positive Ratio (TPR) is sensitivity, Positive Predictive Value (PPV) is precision, and F1 is the harmonic mean between TPR and PPV.

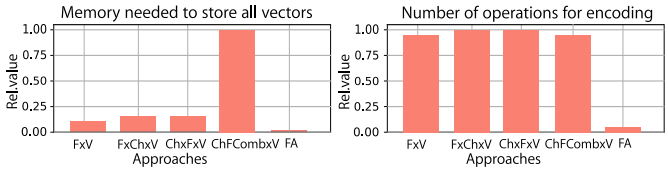

Fig. 6. Comparison of encoding approaches in terms of required memory and number of binding and bundling operations to encode one data window.

From other approaches, $ChFComb \times V$ and $F \times V$ require slightly less operations than $F \times Ch \times V$ and $Ch \times F \times V$.

C. Feature selection

Fig. 7 shows the performance with the FA approach, when features are incrementally added according to the ordering of the three methods described in Sect. 7. In the first column of Fig. 7, features are selected by performance (SBP) (in this case, their F1DE performance), in the second column, they are selected based on average feature confidence (SBC), while in the third column, features are added one-by-one (OBO), selecting the feature that best improves performance when added to previously chosen features. This approach results in a more optimal ordering and choice of features, as it takes into account the correlation between features, and indeed attains the highest performance of the three strategies.

The last column shows the boxplots of feature orders for each feature selection method over all subjects. The smaller the ranking, the sooner that feature was chosen. It can be seen that the classification of features in the first two approaches, using only feature performance or confidence, gives similar results to the discriminative power analysis shown in Fig. 3. In the last case of OBO feature selection, the feature order is slightly different due to the accounting for feature correlations.

Fig. 8 shows the results of OBO feature selection for every subject. The first graph shows the number of features per subject and average feature count for all subjects (the horizontal line). The next two graphs show the performance improvement (or decrease), i.e., F1 for episodes and gmean of F1 for episodes and duration, for the training and test set. The figures' titles contain the average performance over all subjects on the test set. There is significant variability between the number of features chosen per subject, ranging from 1 to 10, with an average of 5.8, or 30%, of the features used.

Table V-C shows the results for the three feature selection methods, when the F1 score for episodes (F1E), or F1 score

Feat. sel. approach	Target Metric	Num. Feat.	Train [%] F1E	Train [%] F1DE	Test [%] F1E	Test [%] F1DE
SBP	F1E	2.84	12.29	4.63	4.78	-1.40
	F1DE	6.99	9.41	7.49	4.94	3.69
SBC	F1E	5.98	10.92	8.55	4.00	3.75
	F1DE	6.62	9.44	10.26	3.56	6.03
OBO	F1E	2.65	14.61	8.91	5.84	1.88
	F1DE	5.81	13.29	10.84	2.45	6.54

TABLE I
OPTIMAL NUMBER OF FEATURES AND PERFORMANCE IMPROVEMENT FOR DIFFERENT FEATURE SELECTION APPROACHES.

for episodes and duration (F1DE) are taken into account. It can be noticed that when optimizing only for F1E, fewer features are needed, but this usually results in a smaller performance increase for F1DE. When F1DE is optimized, this leads to a significant performance increase both for F1DE and F1E but at the price of a slightly higher number of features. In general, the performance increase is smaller on the test set than on the train set, which is reasonable, as the optimal number of features was chosen based on the train set without knowledge about the test set. Yet, the performance increase due to feature selection is still significant, ranging up to 7% for the test set. Overall, SBP and SBC feature selection leads to a slightly higher number of features but not to an improved performance. This is due to the lack of feature correlation information. Finally, the code and data required to reproduce the presented results are available online as open-source[1].

VI. DISCUSSION

This paper draws attention to the topics of mapping and encoding spatio-temporal data such as EEG or EMG to HD vectors. Although HD computing has been utilized and shows promising results for various biomedical applications, most works in the literature use only raw data or LBP values, namely, only one feature per channel. Thus, optimal encoding when more features per channel or more data modalities are used remains unclear. In this work, we propose five strategies to encode feature values as well as channel information into an HD vector and test it on epileptic seizure detection. Our results show that including channel information is beneficial for detection performance, but that the order in which features and channels are mapped is not relevant. Further, we show that the $ChFComb \times V$ approach is the most memory demanding and that $F \times Ch \times V$, $ChxF \times V$ and FA are comparable and more appropriate. FA requires the least amount of memory and operations to encode vectors due to the smaller effective number of dimensions per feature vector, but also results in slightly lower performance.

Using the FA approach, to the best of our knowledge, we present the first method in the literature for performing feature selection using HD computing. An incremental feature selection approach was tested with three different methods to determine the order of features to be added. All approaches led to a significant reduction of features while keeping or even

[1]https://c4science.ch/source/FeatureSelectionWithHD/

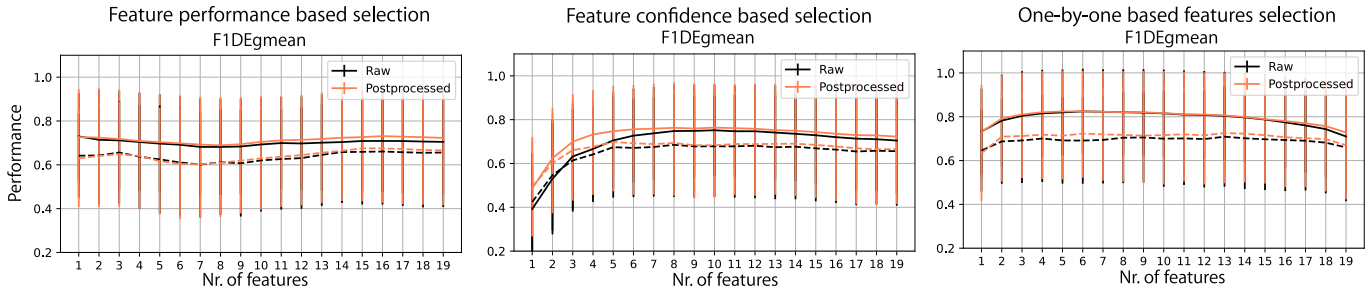

Fig. 7. Performance evolution by incrementally adding new features for three approaches as described in Sect. III-B: SBP, SBC, and OBO.

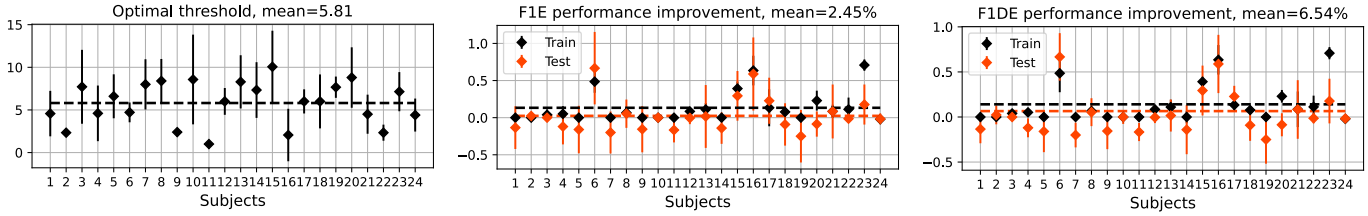

Fig. 8. Optimal number of features and performance after feature selection. Example is shown for feature selection using both feature performance and correlation. The horizontal lines represent the average for all subjects.

significantly improving performance compared to using all features. FA is interesting not only from a feature selection perspective, but also from a clinical perspective, as it leads to a deeper understanding of various features and their properties. For example, we investigated several measures per feature: performance, probabilities of decisions, confidence, correlation, and separability of classes, which can all lead to knowledge discovery related to the usefulness of features.

This approach can be adapted to ChA and can be used analogously for channel comparison, channel selection, and potentially seizure localization. Therefore, we hope that in the future this work will serve as inspiration for further research and novel ideas in the direction of feature exploration, feature and prediction interpretability, and channel selection. More specifically, for the detection of epileptic seizures, we believe that seizure localization and a more detailed feature quality assessment for different seizure-type classifications could be interesting research venues.

VII. Conclusion

In this paper, we have demonstrated how a novel hyperdimensional computing approach can be used as an alternative to state-of-the-art ML in the use case of epileptic seizure detection. We explored the not yet addressed topic of optimal encoding of spatio-temporal ExG data, such as EEG, and all the information it entails, into HD vectors. We compared five different approaches with respect to their memory and computational complexity, as these metrics are of great interest for wearable devices for continuous monitoring of diseases. Hence, those approaches with lower complexity could make longer battery lifetimes feasible and make a step towards preventive healthcare. In addition, we have demonstrated how the HD computing framework can be utilized to perform feature selection by choosing a proper encoding strategy. Three approaches were tested and led to an average feature reduction of 70%, while maintaining or even significantly improving detection accuracy (up to 7%) compared to using all the features. We trust that this work can serve as inspiration for further research and novel ideas in the direction of feature exploration, feature and prediction interpretability, and channel selection.

References

[1] G. Karunaratne *et al.*, "Energy Efficient In-Memory Hyperdimensional Encoding for Spatio-Temporal Signal Processing," *IEEE TCAS*, 2021.

[2] A. Burrello *et al.*, "An Ensemble of Hyperdimensional Classifiers: Hardware-Friendly Short-Latency Seizure Detection With Automatic iEEG Electrode Selection," *IEEE JBHI*, 2021.

[3] F. Asgarinejad *et al.*, "Detection of Epileptic Seizures from Surface EEG Using Hyperdimensional Computing," in *IEEE EMBC*, 2020.

[4] A. Burrello *et al.*, "Laelaps: An Energy-Efficient Seizure Detection Algorithm from Long-term Human iEEG Recordings without False Alarms," *DATE*, 2019.

[5] Y. Kaya *et al.*, "1D-local binary pattern based feature extraction for classification of epileptic EEG signals," *Appl. Math. Comput.*, 2014.

[6] U. Pale *et al.*, "Systematic Assessment of Hyperdimensional Computing for Epileptic Seizure Detection," *IEEE EMBC*, 2021.

[7] U. Pale *et al.*, "Exploration of Hyperdimensional Computing Strategies for Enhanced Learning on Epileptic Seizure Detection," *arXiv:2201.09759*, 2022.

[8] S. Benatti *et al.*, "Online Learning and Classification of EMG-Based Gestures on a Parallel Ultra-Low Power Platform Using Hyperdimensional Computing," *IEEE TBioCAS*, 2019.

[9] L. Ge and K. K. Parhi, "Seizure Detection Using Power Spectral Density via Hyperdimensional Computing," in *IEEE ICASSP*, 2021.

[10] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Thesis, MIT, 2009.

[11] U. Pale *et al.*, "Multi-Centroid Hyperdimensional Computing Approach for Epileptic Seizure Detection," *Frontiers in Neurology*, 2022.

[12] R. Esteller *et al.*, "Line length: An efficient feature for seizure onset detection," in *IEEE EMBC*, 2001.

[13] A. Hernández-Cano *et al.*, "A Framework for Efficient and Binary Clustering in High-Dimensional Space," in *DATE*, 2021.